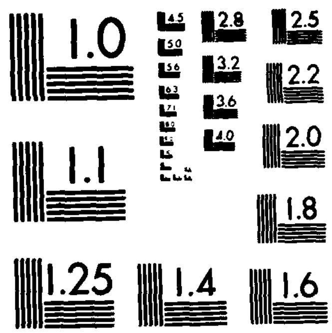
AD-A192 716 CONTENT-ADDRESSABLE MEMORY STORAGE BY NEURAL NETWORKS: 1/1
A GENERAL MODEL AN.. (U) BOSTON UNIV MA CENTER FOR
ADAPTIVE SYSTEMS S GROSSBERG MAR 88 AFOSR-TR-88-0431
UNCLASSIFIED F49620-86-C-0037 F/G 12/9 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A192 716

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 88-0431	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER DTIC FILE COPY
4. TITLE (and Subtitle) Content-Addressable Memory Storage By Neural Networks: A General Model And Global Liapunov Method		5. TYPE OF REPORT & PERIOD COVERED ~~Interim report~~ Reprint
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen Grossberg		8. CONTRACT OR GRANT NUMBER(s) ~~AFOSR~~ F49620-86-C-0037 AF~~OSR~~ F49620-87-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Adaptive Systems Department of Mathematice, Boston University Boston, MA 02215		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A7
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR Life Sciences Directorate Bolling Air Force Base Washington, DC 20332 NL		12. REPORT DATE March 1988
		13. NUMBER OF PAGES 22 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same as 11		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited

DTIC ELECTE MAY 02 1988 S D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

E

18. SUPPLEMENTARY NOTES

To appear in E.L. Schwartz (Ed.), Computational Neuroscience Cambridge, MA: MIT Press, 1988

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Many neural network models capable of content-addressable memory are shown to be special cases of the general model and global Liapunov function described by Cohen and Grossberg (1983). These include examples of the additive, brain-state-in-a-box, McCulloch-Pitts, Boltzmann machine, shunting, masking field, bidirectional associative memory, Volterra-Lotka, Gilpin-Ayala, and Eigen-Schuster models. The Cohen-Grossberg model thus defines a general principle for the design of content-addressable memory, that is shared by all model

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

20. Abstract (continued)

exemplars of such a general design constitutes a computational invariant. Such a general model and analytic method defines a computational framework within which specialized model exemplars may be compared to discover which models are best able to explain particular parametric data about brain and behavior, or to solve particular technological problems.

AFOSR-TR- 88-0431

CONTENT-ADDRESSABLE MEMORY STORAGE BY NEURAL NETWORKS: A GENERAL MODEL AND GLOBAL LIAPUNOV METHOD

Stephen Grossberg†
Center for Adaptive Systems
Boston University
111 Cummington Street
Boston, MA 02215

To appear in E.L. Schwartz (Ed.), Computational Neuroscience
Cambridge, MA: MIT Press, 1987

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A-1	

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-86-C-0037 and ~~AFOS~~R F49620-87-C-0018) and the National Science Foundation (NSF IRI-84-17756).

Acknowledgements: Thanks to Cynthia Suchta and Carol Yanakakis for their valuable assistance in the preparation of the manuscript and illustrations.

88 5 02 137

ABSTRACT

Many neural network models capable of content-addressable memory are shown to be special cases of the general model and global Liapunov function described by Cohen and Grossberg (1983). These include examples of the additive, brain-state-in-a-box, McCulloch-Pitts, Boltzmann machine, shunting, masking field, bidirectional associative memory, Volterra-Lotka, Gilpin-Ayala, and Eigen-Schuster models. The Cohen-Grossberg model thus defines a general principle for the design of content-addressable networks. A model-independent property, such as content-addressable memory, that is shared by all model exemplars of such a general design constitutes a computational invariant. Such a general model and analytic method defines a computational framework within which specialized model exemplars may be compared to discover which models are best able to explain particular parametric data about brain and behavior, or to solve particular technological problems.

1. The Role of Design Principles and Computational Invariants in Computational Neuroscience

Due to the highly interactive nature of brain dynamics, the analysis of general design principles, cellular mechanisms, network modules, and functionally specialized multi-level architectures have proceeded hand-in-hand, with each level of analysis clarifying the scientific understanding of the other levels. In all of these modelling endeavors, mathematical analysis and parametric computer simulations have played an important role by demonstrating how complex emergent properties may arise from interactions among simpler network components. Mathematical analysis has also been particularly useful towards the identification of general design principles by clarifying how ostensibly different specialized model mechanisms may all be variations of a general computational theme, in that they all exhibit one or more important computational invariants.

The mathematical characterization of such a computational invariant greatly facilitates a finer, comparative analysis which cuts across the specialized models that are capable of generating the invariant. By sharply distinguishing the general from the particular, such a comparative analysis of specialized models helps to discover which model is best able to explain particular parametric behavior about brain and behavior, or to solve a particular technological problem. Thus, after such a general design principle or computational invariant is discovered, representative models may be reorganized in a way that pays less attention to accidents of their historical discovery and development than to their functional and computational properties. The identification of such design principles and computational invariants is thus an important goal of computational neuroscience.

2. Content-Addressable Memory and Liapunov Methods

The present chapter describes a general design principle for constructing neural networks that are capable of content-addressable memory, or CAM. From a mathematical perspective, the question of content-addressable memory in a neural network can be for-

mulated as follows: Under what conditions does a neural network always approach an equilibrium point in response to an arbitrary, but sustained, input pattern? The equilibrium point represents the stored pattern in response to the input pattern. In a satisfactory analysis of this problem, the behavior of the network in response to arbitrary initial data, an arbitrary sustained input pattern, and an arbitrary choice of network parameters is provided. Also an account of how many equilibrium points exist and of how they are approached through time is desirable. Such a mathematical analysis is called a global analysis, to distinguish it from a local stability analysis around individual equilibrium points.

Amari and Arbib (1982) and Levine (1983) have described a number of contributions to the local analysis of neural networks. Our concern herein is with global methods. A global analysis of equilibrium behavior is of importance for an understanding both of CAM and of the types of nonequilibrium behavior—such as travelling waves, bursts, standing waves, and chaos which can be obtained by perturbing off systems which always approach equilibrium (Carpenter, 1977a, 1977b, 1979, 1981; Cohen and Grossberg, 1983; Ellias and Grossberg, 1975: Ermentrout and Cowan, 1979, 1980; Hastings, 1976, 1982; Hodgson, 1983; Kaczmarek and Babloyantz, 1977). A global mathematical analysis of nonlinear associative learning networks was begun in Grossberg (1967, 1968b). A global mathematical analysis of nonlinear shunting cooperative-competitive feedback networks was begun in Grossberg (1973). Some of the main articles in these series are brought together in Grossberg (1982).

One technique for the global approach to equilibrium which has attracted widespread interest is the use of global Liapunov, or energy, methods. Global Liapunov methods were introduced for the analysis of increasingly large classes of neural networks in the 1970's (Grossberg, 1977, 1978a, 1978b, 1980, 1981).

Prior to the use of Liapunov methods for the analysis of neural networks, such methods were used to analyse other biological networks, notably networks arising in mathematical ecology. For example, MacArthur (1970) described a quadratic Liapunov function for

proving local asymptotic stability of isolated equilibrium points of Volterra-Lotka systems with symmetric coefficients. Goh and Agnew (1977) described a global Liapunov function for Volterra-Lotka and Gilpin-Ayala systems in cases where only one equilibrium point exists. Liapunov functions were also described for Volterra-Lotka systems whose off-diagonal terms are relatively small (Kilmer, 1972; Takeuchi, Adachi, and Tokumaru, 1978). Such constraints were, however, too limiting for the design of CAM systems aimed at transforming and storing a large variety of patterns. Herein I summarize a general model of a nonlinear cooperative-competitive neural network for which a global Liapunov function was explicitly constructed to ensure CAM. I then show that a number of popular models are special cases of the general model, and thus are capable of CAM.

Cohen and Grossberg (1983) described a general principle for designing CAM networks by proving that models that can be written in the form

$$\frac{d}{dt} x_i = a_i(x_i)[b_i(x_i) - \sum_{j=1}^{n} c_{ij} d_j(x_j)] \tag{1}$$

admit the global Liapunov function

$$V = -\sum_{i=1}^{n} \int^{x_i} b_i(\xi_i) d_i'(\xi_i) d\xi_i + \frac{1}{2} \sum_{j,k=1}^{n} c_{jk} d_j(x_j) d_k(x_k) \tag{2}$$

if the coefficient matrix $C = \| c_{ij} \|$ and the functions a_i, b_i, and d_j obey mild technical conditions, including

Symmetry:

$$c_{ij} = c_{ji}. \tag{3}$$

Positivity:

$$a_i(x_i) \geq 0 \tag{4}$$

Monotonicity:

$$d_j'(x_j) \geq 0. \tag{5}$$

Integrating V along trajectories implies that

$$\frac{d}{dt}V = -\sum_{i=1}^{n} a_i d_i' [b_i - \sum_{j=1}^{n} c_{ij} d_j]^2. \tag{6}$$

If (4) and (5) hold, then $\frac{d}{dt}V \leq 0$ along trajectories. Once this basic property of a Liapunov function is in place, it is a technical matter to rigorously prove that every trajectory approaches one of a possibly large number of equilibrium points.

For expository vividness, the functions in the Cohen-Grossberg model (1) are called the *amplification* function a_i, the *self-signal* function b_i, and the *other-signal* functions d_j. Specialized models are characterized by particular choices of these functions.

3. Additive Model

The systematic physical and mathematical development of the additive model began in the late 1960's; e.g., in Grossberg (1967, 1968a, 1968b, 1969a, 1969b, 1970a, 1970b) and Grossberg and Pepe (1971). Cohen and Grossberg (1983, p.819) noted that "the simpler additive neural networks ... are also included in our analysis". The additive equation can be written using the coefficients of the standard electrical circuit interpretation (Plonsey and Fleming, 1969) as

$$C_i \frac{dx_i}{dt} = -\frac{1}{R_i} x_i + \sum_{j=1}^{n} f_j(x_j) z_{ji} + I_i. \tag{7}$$

Substitution into (1) shows that

$$a_i(x_i) = \frac{1}{C_i} \quad \text{(constant!)} \tag{8}$$

$$b_i(x_i) = \frac{1}{R_i} x_i + I_i \quad \text{(linear!)} \tag{9}$$

$$c_{ij} = -T_{ij} \tag{10}$$

and

$$d_j(x_j) = f_j(x_j). \tag{11}$$

Thus in the additive case, the amplification function (8) is a positive constant, hence satisfies (4), and the self-signal term (9) is linear. Substitution of (8)-(11) into (2) leads directly to the equation

$$V = \sum_{i=1}^{n} \frac{1}{R_i} \int^{x_i} \xi_i f_i'(\xi_i) d\xi_i - \sum_{i=1}^{n} I_i f_i(x_i) - \frac{1}{2} \sum_{j,k=1}^{n} T_{jk} f_j(x_j) f_k(x_k). \tag{12}$$

This Liapunov function for the additive model was later published by Hopfield (1984). In Hopfield's treatment, ξ_i is written as an inverse $f_i^{-1}(V_i)$. Cohen and Grossberg (1983) showed, however, that although $f_i(x_i)$ must be nondecreasing, as in (5), it need not have an inverse in order for (12) to be valid.

4. Brain-State-in-a-Box Model: $S\sum$ Exchange

The BSB model was introduced in Anderson, Silverstein, Ritz, and Jones (1977). It is often described in discrete time by the equation

$$x_i(t+1) = S(x_i(t) + \alpha \sum_{j=1}^{n} A_{ij} x_j(t)) \tag{13}$$

using symmetric coefficients

$$A_{ij} = A_{ji} \tag{14}$$

and a special type of nonlinear signal function $S(w)$ that characterizes the model. The signal function is a symmetric ramp function:

$$S(w) = \begin{cases} F & \text{if } w \geq F \\ w & \text{if } -F < w < F \\ -F & \text{if } w \leq -F \end{cases} . \tag{15}$$

Thus each STM trace x_i obeys a *linear* equation until its argument reaches the *hard saturation* limit F.

The BSB model has been used to discuss categorical perception in terms of its formal contrast enhancement property that each x_i tends to approach a limiting value $\pm F$, and thus that the vector $(x_1, x_2, \ldots, x_n)$ tends to approach a corner of the box $(\pm F, \pm F, \ldots, \pm F)$

as time goes on. An alternative explanation of contrast enhancement by a nonlinear feedback network was provided in Grossberg (1973) using a sigmoid signal function, rather than a function linear near zero, coupled to the soft saturation dynamics of a shunting network, rather than the hard saturation of a symmetric ramp. This is still a topic undergoing theoretical discussion (Anderson, Silverstein, Ritz, and Jones, 1977; Grossberg, 1978c, 1987d).

The BSB model can be rewritten as an additive model with no input and a special signal function that satisfies (5). Hence it is a special case of model (1). To see this, rewrite (13) in the form

$$x_i(t+1) = S\left(\sum_{j=1}^{n} B_{ij} x_j(t)\right) \tag{16}$$

using the coefficient

$$B_{ij} = \delta_{ij} + \alpha A_{ij} \tag{17}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$. By (14), it follows that

$$B_{ij} = B_{ji}. \tag{18}$$

Although (16) is written in discrete time for computational convenience, it needs to be expressed in continuous time in order to represent a physical model, as in

$$\frac{d}{dt} x_i = -x_i + S\left(\sum_{j=1}^{n} B_{ij} x_j\right). \tag{19}$$

Define the new variables y_i by

$$y_i = \sum_{j=1}^{n} B_{ij} x_j. \tag{20}$$

Then

$$\frac{d}{dt} y_i = -y_i + \sum_{j=1}^{n} B_{ij} S(y_j). \tag{21}$$

Comparison of (21) with (7) shows that the BSB model is an additive model such that each input $I_i = 0$. This simple change of coordinates from (19) to (21) is so important in neural modelling that I give it a name: *Signal-Sum ($S\sum$) Exchange*.

The observation that, via $S\sum$ Exchange, a nonlinear signal of a sum, as in (19), can be rewritten as a sum of nonlinear signals, as in (21), shows that a number of models which have been treated as distinct are mathematically identical. In contrast, this type of transformation cannot be carried out on shunting models such as (31) below.

The Liapunov function for (21) is found by directly substituting into model (1) expressed in terms of the variables y_i:

$$\frac{d}{dt}y_i = a_i(y_i)[b_i(y_i) - \sum_{j=1}^{n} c_{ij}d_j(y_j)]. \tag{22}$$

Since $a_i(y_i) = 1$, $b_i(y_i) = -y_i$, $c_{ij} = -B_{ij}$, and $d_j(y_j) = S(y_j)$, substitution into (2) yields

$$V = \sum_{i=1}^{n} \int^{y_i} \xi_i S'(\xi_i)d\xi_i - \frac{1}{2}\sum_{k=1}^{n} B_{jk}S(y_j)S(y_k). \tag{23}$$

Using the definitions in (15), (17), and (20), (23) can be rewritten in terms of the original variables x_i as follows:

$$V = -\frac{\alpha}{2}\sum_{j,k=1}^{n} A_{jk}x_jx_k. \tag{24}$$

Golden (1986) has derived (24) from a direct analysis of the BSB model.

5. The McCulloch-Pitts Model

This classical model takes the form

$$x_i(t+1) = \text{sgn}(\sum_{j=1}^{n} A_{ij}x_j(t) - B_i). \tag{25}$$

Letting

$$M(w) = \text{sgn}(w - B_i), \tag{26}$$

(25) can be rewritten as

$$x_i(t+1) = M(\sum_{j=1}^{n} A_{ij}x_j(t)). \tag{27}$$

As in the analysis of (19), (27) can be rewritten in continuous time in terms of the variables y_i via $S\sum$ Exchange:

$$\frac{d}{dt}y_i = -y_i + \sum_{j=1}^{n} A_{ij}M(y_j) \tag{28}$$

and is thus also a symmetric additive model with zero inputs. In addition, its signal function $M(y_j)$ has a zero derivative ($M'(y_j) = 0$) except at $y_j = 0$. Substitution of this additional property into (23) shows that the Liapunov function for the continuous time McCulloch-Pitts model is

$$V = -\frac{1}{2} \sum_{j,k=1}^{n} A_{jk} M(y_j) M(y_k), \tag{29}$$

which is the continuous time version of the discrete time Liapunov function described by Hopfield (1982).

6. The Boltzmann Machine

The state equation of the Boltzmann machine (Ackley, Hinton, and Sejnowski, 1985) is also an additive equation with symmetric coefficients. Its signal function is the sigmoid logistic function

$$f(w) = \frac{1}{1 + e^{-w}}, \tag{30}$$

which satisfies (5) and is thus a special case of model (1). Thus the Boltzmann machine is a specialized additive model regulated by simulated annealing (Geman, 1983, 1984; Kirkpatrick, Gelatt, and Vecchi, 1982).

7. Shunting Cooperative-Competitive Feedback Network

All additive models lead to constant amplification functions $a_i(x_i)$ and linear self-feedback functions $b_i(x_i)$. The need for the more general model (1) becomes apparent when the shunting model is analysed. Consider, for example, a class of shunting models in which each node can receive excitatory and inhibitory inputs I_i and J_i, respectively, and each node can excite itself and can inhibit other nodes via nonlinear feedback. Such networks model on-center off-surround interactions among cells which obey membrane equations (Grossberg, 1973; Hodgkin, 1964; Kandel and Schwartz, 1981; Katz, 1966; Plonsey and Fleming, 1969). In particular, let

$$\frac{d}{dt} x_i = -A_i x_i + (B_i - x_i)[I_i + f_i(x_i)] - (x_i + C_i)[J_i + \sum_{j=1}^{n} D_{ij} g_j(x_j)]. \tag{31}$$

In (31), each x_i can fluctuate within the finite interval $[-C_i, B_i]$ in response to the constant inputs I_i and J_i, the state-dependent positive feedback signal $f_i(x_i)$, and the negative feedback signals $D_{ij}g_j(x_j)$. It is assumed that

$$D_{ij} = D_{ji} \geq 0 \tag{32}$$

and that

$$g_j'(x_j) \geq 0. \tag{33}$$

In order to write (31) in Cohen-Grossberg form, it is convenient to introduce the variables

$$y_i = x_i + C_i. \tag{34}$$

In applications, C_i is typically nonnegative. Since x_i can vary within the interval $[-C_i, B_i]$, y_i can vary within the interval $[0, B_i + C_i]$ of nonnegative numbers. In terms of these variables, (31) can be written in the form

$$\frac{d}{dt} y_i = a_i(y_i)[b_i(y_i) - \sum_{j=1}^{n} C_{ij} d_j(y_j)] \tag{22}$$

where

$$a_i(y_i) = y_i \quad \text{(nonconstant!)}, \tag{35}$$

$$b_i(y_i) = \frac{1}{y_i}[A_i C_i - (A_i + J_i)y_i + (B_i + C_i - y_i)(I_i + f_i(y_i - C_i))] \quad \text{(nonlinear!)}, \tag{36}$$

$$C_{ij} = D_{ij}, \tag{37}$$

and

$$d_j(y_j) = g_j(y_j - C_j) \quad \text{(noninvertible!)}. \tag{38}$$

Unlike the additive model, the amplification function $a_i(y_i)$ in (35) is not a constant. In addition, the self-signal function $b_i(y_i)$ in (36) is not necessarily linear, notably because the feedback signal $f_i(x_i - C_i)$ is often nonlinear in applications of the shunting model; in particular it is often a sigmoid or multiple sigmoid signal function (Ellias and Grossberg,

1975; Grossberg, 1973, 1977, 1978b; Grossberg and Levine, 1975; Sperling, 1981). Sigmoid signal functions, and approximations thereto, also appear in applications of the additive model and its variants (Ackley, Hinton, and Sejnowski, 1985; Amari and Arbib, 1982; Freeman, 1975, 1979; Grossberg, 1969a, 1982; Grossberg and Kuperstein, 1986; Hinton and Anderson, 1981; Hopfield, 1984; Rumelhart and McClelland, 1986). Such applications do not require the full generality of the Liapunov function (1) because the nonlinear signal function can then be absorbed into the terms $d_j(x_j)$.

Property (4) follows from the fact that $a_i(y_i) = y_i \geq 0$. Property (5) follows from the assumption that the negative feedback signal function g_j in (38) is monotone nondecreasing. Cohen and Grossberg (1983) proved that g_j need not be invertible. A signal threshold may exist below which $g_j = 0$ and above which g_j may grow in a nonlinear way. The inclusion of nonlinear signals with thresholds better enables the model to deal with fluctuations due to subthreshold noise. On the other hand, thresholds are not the only mechanisms which can suppress noise in a cooperative-competitive feedback network.

8. Masking Field Model

In many applications of the shunting and additive models, the coefficients c_{ij} in (1) may be asymmetric, thereby rendering the Liapunov function (2) inapplicable. Asymmetric coefficients typically occur in problems relating to the learning and recognition of temporal order in behavior. Consequently, a number of mathematical methods have been developed from the earliest days of neural network theory to analyse models with asymmetric interaction coefficients.

On the other hand, certain network models may have asymmetric interaction coefficients, yet be reduceable to the form (1) with symmetric interaction coefficients through a suitable change of variables. The masking field model is a shunting network of this type. The masking field model was introduced in Grossberg (1978d; reprinted in Grossberg, 1982) to explain data about speech learning, word recognition, and the learning of adap-

tive sensory-motor plans. It has been further developed through computer simulations in Cohen and Grossberg (1986, 1987). A masking field is a multiple-scale, self-similar, automatically gain controlled, cooperative-competitive nonlinear feedback network which can generate a compressed but distributed STM representation of an input pattern as a whole, of its most salient parts, and of predictive codes which represent larger input patterns of which it forms a part. The masking field model is thus a specialized type of vector quantization scheme (Gray, 1984). Its multiple-scale self-similar properties imply its asymmetric interaction coefficients.

The STM equation of a typical masking field is defined by

$$\frac{d}{dt}x_i^{(J)} = -Ax_i^{(J)} + (B - x_i^{(J)})[\sum_{j\in J} E_j p_{ji}^{(J)} + D \mid J \mid f(x_i^{(J)})] \\ - (x_i^{(J)} + C)\frac{\sum_{m,K} g(x_m^{(K)}) \mid K \mid (1+ \mid K \cap J \mid)}{\sum_{m,K} \mid K \mid (1+ \mid K \cap J \mid)}. \tag{39}$$

In (39), $x_i^{(J)}$ is the STM trace of the ith masking field node that receives excitatory input $\sum_{j\in J} E_j p_{ji}^{(J)}$ from the unordered set J of input items. Notation $\mid J \mid$ counts the number of items in set J and thereby keeps track of the number of spatial scales that go into each version of the model.

The inhibitory interaction coefficient

$$\frac{\mid K \mid (1+ \mid K \cap J \mid)}{\sum_{m,K} \mid K \mid (1+ \mid K \cap J \mid)} \tag{40}$$

in (39) is an asymmetric function of J and K. Despite this fact, (39) can be written in Cohen-Grossberg form as

$$\frac{d}{dt}y_i^{(J)} = a_i^{(J)}(y_i^{(J)})[b_i^{(J)}(y_i^{(J)}) - \sum_{m,K} c_{KJ} d^{(K)}(y_m^{(K)})] \tag{41}$$

with symmetric coefficients

$$c_{KJ} = c_{JK} = 1+ \mid K \cap J \mid \tag{42}$$

in terms of the variables

$$y_i^{(J)} = F_{|J|}^{-1}(x_i^{(J)} + C) \tag{43}$$

where

$$F_{|J|} = \sum_{m,K} | K | (1+ | K \cap J |). \tag{44}$$

This is seen as follows. Since $F_{|J|}$ is the denominator of (40), it can be used to divide term $x_i^{(J)} + C$ in (39). Then the asymmetric term $| K |$ in the numerator of (40) can be absorbed into the definition of g in (39). Then by redefining and rearranging terms as in (35)-(38), equation (41) holds with

$$a_i^{(J)}(y_i^{(J)}) = F_{|J|}^{-1} y_i^{(J)} \tag{45}$$

$$\begin{aligned} b_i^{(J)}(y_i^{(J)}) = \frac{1}{y_i^{(J)}}[&AC - AF_{|J|}y_i^{(J)} + \left(\frac{B+C}{F_{|J|}} - y_i^{(J)}\right) \\ &(I_i^{(J)} + D \mid J \mid F_{|J|} f(F_{|J|} y_i^{(J)} - C))] \end{aligned} \tag{46}$$

where

$$I_i^{(J)} = F_{|J|} \sum_{j \in J} E_j p_{ji}^{(J)}, \tag{47}$$

and

$$d^{(K)}(y_m^{(K)}) = | K | g(F_{|K|} y_m^{(K)} - C). \tag{48}$$

Thus the masking field model is a specialized Cohen-Grossberg model.

9. Bidirectional Associative Memories: Symmetrizing an Asymmetric Interaction Matrix

Other procedures have also been devised for dealing with systems having asymmetric coefficients; for example, given an arbitrary $n \times m$ coefficient matrix $Z = \| z_{ij} \|$ from a network level F_1 to a network level F_2 with STM traces x_i and y_j, respectively. Kosko and Guest (1987) and Kosko (1987) have shown that (1) and (2) can be used to construct feedback pathways from F_2 to F_1 so that the two-level feedback network $F_1 \leftarrow F_2$ has convergent trajectories.

For example, if the bottom-up interaction $F_1 \to F_2$ obeys an additive equation

$$\frac{d}{dt} y_j = -A_j y_j + \sum_k f_k(x_k) z_{kj} + I_j, \tag{49}$$

then the top-down interaction $F_2 \to F_1$ is defined to obey an additive equation

$$\frac{d}{dt} x_i = -B_i x_i + \sum_l g_l(y_l) z_{il} + J_i, \tag{50}$$

where I_j and J_i are input terms. This definition creates a symmetric interaction matrix by *closing the top-down feedback loop, since if* $f_i(x_i)$ *influences* y_j with coefficient z_{ij} in (49), then $g_j(y_j)$ influences x_i with the same coefficient z_{ij}. Thus by defining an augmented vector $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ of STM activities, system (49)–(50) as a whole defines an additive model (7) with an $(n+m) \times (n+m)$ symmetric coefficient matrix.

The same procedure can be used to symmetrize many other neural network models. Kosko and Guest (1987) have described optical implementations for this procedure, and Kosko (1987) has used the symmetrized additive model to discuss minimization of fuzzy entropy.

10. Volterra-Lotka, Gilpin-Ayala, and Eigen-Schuster Models

The Cohen-Grossberg model was designed to also include models which arose in other areas of biology than neural network theory. For example, it includes the classical

Volterra-Lotka Model

$$\frac{d}{dt} x_i = A_i x_i (1 - \sum_{j=1}^{n} B_{ij} x_j) \tag{51}$$

of population biology (Lotka, 1956), the

Gilpin-Ayala Model

$$\frac{d}{dt} x_i = A_i x_i \left[1 - \left(\frac{x_i}{B_i}\right)^{\theta_i} - \sum_{j=1}^{n} C_{ij} \left(\frac{x_j}{B_j}\right)\right], \tag{52}$$

also from population biology (Gilpin and Ayala, 1973), and the

Eigen-Schuster Model

$$\frac{d}{dt}x_i = x_i(A_i x_i^{p-1} - q\sum_{j=1}^{n} A_j x_j^p) \tag{53}$$

from the theory of macromolecular evolution (Eigen and Schuster, 1978). In all of these models, either the amplification function $a_i(x_i)$ is non-constant, or the self-signal function $b_i(x_i)$ is nonlinear, or both.

11. Concluding Remarks: Comparative Analysis and Model Selection

The specialized models summaried in Sections 2-9 illustrate that model (1) and Liapunov function (2) embody a general principle for designing CAM devices from cooperative-competitive feedback models. These models are said to be *absolutely stable* because the CAM property is not destroyed by changing the parameters, inputs, or initial values of the model. The persistence of the CAM property under arbitrary parameter changes enables learning to change system parameters in response to unpredictable input environments without destroying CAM. Nonetheless, the transformation from input patterns to stored patterns executed by a network with adaptively altered parameters can differ significantly from its original transformation.

The Cohen-Grossberg analysis emphasizes the critical role of mathematical methods in classifying and understanding very large systems of nonlinear neural networks (VLSN). Without such an integrative approach, it is sometimes difficult to tell even whether or not a model is really new computationally, or whether it is a variant of a known model in different coordinates or notation. For example, many scientists have not realized that models (19) and (21) are mathematically equivalent. Table 1 describes the relationships between models summarized herein by such an analysis. Thus the Brain-State-in-a-Box and Boltzmann machine models enjoy a CAM property for the same reason that any additive or Cohen-Grossberg model does. On the other hand, the BSB model may have special properties that may make it ideal for certain tasks, or it may be too specialized to

accomplish certain tasks which are better dealt with using a more general additive model, or even a shunting model.

Table 1

The set of specialized properties which differentiate one model from another ultimately provides the computational rationale for continued interest in that model, whether because its set of properties better explain an important behavioral or neural data base, or because these properties enable the model to more efficiently solve an important technological problem. Indeed, the Cohen-Grossberg (1983) model was developed to integrate many of the specialized additive and shunting models which had arisen in applications since the late 1960's. Once such an integrative design is recognized, a comparative analysis of specialized models is mandated within the computational framework defined by the design principle and its unifying mathematical method. Such a comparative analysis needs to clearly distinguish between the model-independent properties that are shared by all models which exemplify the design and the unique combinations of model-dependent properties which both the evolutionary process and the human engineer seek to utilize to their maximal advantage.

TABLE 1

CG (1983)	ADDITIVE (1967)	MP (1943)
		BSB (1977)
		BM (1985)
		BAM (1987)
	SHUNTING (1973)	MF (1978, 1986)

Organization in terms of decreasing generality of the models described in Section 9. Abbreviations: CG = Cohen-Grossberg; MP = McCulloch-Pitts; BSB = Brain-State-in-a-Box: BM = Boltzmann Machine; BAM = Bidirectional Associative Memory; MF = Masking Field.

References

Ackley, D.H., Hinton, G.E., and Sejnowski, T.J., A learning algorithm for Boltzmann machines. *Cognitive Science*, 1985, **9**, 147–169.

Amari, S. and Arbib, M.A. (Eds.); **Competition and cooperation in neural networks**. New York: Springer-Verlag, 1982.

Anderson, J.A., Silverstein, J.W., Ritz, S.R., and Jones, R.S., Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 1977, **84**, 413–451.

Carpenter, G.A., A geometric approach to singular perturbation problems with applications to nerve impulse equations. *Journal of Differential Equations*, 1977, **23**, 335–367 (a).

Carpenter, G.A., Periodic solutions of nerve impulse equations. *Journal of Mathematical Analysis and Applications*, 1977, **58**, 152–173 (b).

Carpenter, G.A., Bursting phenomena in excitable membranes. *SIAM Journal on Applied Mathematics*, 1979, **36**, 334–372.

Carpenter, G.A., Normal and abnormal signal patterns in nerve cells. In S. Grossberg (Ed.), **Mathematical psychology and psychophysiology**. Providence, RI: American Mathematical Society, 1981, 49–90.

Cohen, M.A. and Grossberg, S., Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *Transactions IEEE*, 1983, **SMC-13**, 815–826.

Cohen, M. and Grossberg, S., Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology*, 1986, **5**, 1–22.

Cohen, M. and Grossberg, S., Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, in press, 1987.

Eigen, M. and Schuster, P. The hypercycle: A principle of natural self-organization, B: The abstract hypercycle. *Naturwissenschaften*, 1978, **65**, 7–41.

Ellias, S.A. and Grossberg, S., Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics*, 1975, **20**, 69–98.

Ermentrout, G.B. and Cowan, J.D., Temporal oscillations in neuronal nets. *Journal of Mathematical Biology*, 1979, **7**, 265–286.

Ermentrout, G.B. and Cowan, J.D., Large scale spatially organized activity in neural nets. *SIAM Journal on Applied Mathematics*, 1980, **38**, 1–21.

Freeman, W.J., **Mass action in the nervous system**. New York: Academic Press, 1975.

Freeman, W.J., EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb. *Biological Cybernetics*, 1979, **35**, 221–234.

Geman, S., Stochastic relaxation methods for image restoration and expert systems. *Proceedings of ARO Workshop: Unsupervised Image Analysis*. Brown University, April, 1983.

Geman, S., Stochastic relaxation methods for image restoration and expert systems. In D.B. Cooper, R.L. Launer, and D.E. McClure (Eds.), **Automated image analysis: Theory and experiments**. New York: Academic Press, 1984.

Gilpin, M.E. and Ayala, F.J., Global models of growth and competition. *Proceedings of the National Academy of Sciences*, 1973, **70**, 3590–3593.

Goh, B.S. and Agnew, T.T., Stability in Gilpin and Ayala's models of competition. *Journal of Mathematical Biology*, 1977, **4**, 275–279.

Golden, R.M., The "Brain-state-in-a-box" neural model is a gradient descent algorithm. *Journal of Mathematical Psychology*, 1986, **30**, 73-80.

Gray, R.M., Vector quantization. *IEEE ASSP Magazine*, April, 1984, 4-29.

Grossberg, S., Nonlinear difference-differential equations in prediction and learning theory. *Proceedings of the National Academy of Sciences*, 1967, **58**, 1329-1334.

Grossberg, S., Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, 1968, **60**, 758-765 (a).

Grossberg, S., Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proceedings of the National Academy of Sciences*, 1968, **59**, 368-372 (b).

Grossberg, S., On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, 1969, **1**, 319-350 (a).

Grossberg, S., On learning of spatiotemporal patterns by networks with ordered sensory and motor components, I: Excitatory components of the cerebellum. *Studies in Applied Mathematics*, 1969, **48**, 105-132 (b).

Grossberg, S., Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II. *Studies in Applied Mathematics*, 1970, **49**, 135-166 (a).

Grossberg, S., Neural pattern discrimination. *Journal of Theoretical Biology*, 1970, **27**, 291-337 (b).

Grossberg, S., Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 1973, **52**, 217-257.

Grossberg, S., Pattern formation by the global limits of a nonlinear competitive interaction in *n* dimensions. *Journal of Mathematical Biology*, 1977, **4**, 237-256.

Grossberg, S., Decisions, patterns, and oscillations in the dynamics of competitive systems with applications to Volterra-Lotka systems. *Journal of Theoretical Biology*, 1978, **73**, 101-130 (a).

Grossberg, S., Competition, decision, and consensus. *Journal of Mathematical Analysis and Applications*, 1978, **66**, 470-493 (b).

Grossberg, S., Do all neural models really look alike? *Psychological Review*, 1978, **85**, 592-596 (c).

Grossberg, S., A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), **Progress in theoretical biology**, Vol. 5. New York: Academic Press, 1978 (d).

Grossberg, S., Biological competition: Decision rules, pattern formation, and oscillations. *Proceedings of the National Academy of Sciences*, 1980, **77**, 2338-2342.

Grossberg, S., (Ed.), Adaptive resonance in development, perception, and cognition. In **Mathematical psychology and psychophysiology**. Providence, RI: American Mathematical Society, 1981.

Grossberg, S., **Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control**. Boston: Reidel Press, 1982.

Grossberg, S., Cortical dynamics of three-dimensional form, color, and brightness perception: I: Monocular theory. *Perception and Psychophysics*, 1987, **41**, 87-116 (a).

Grossberg, S., Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory. *Perception and Psychophysics*, 1987, **41**, 117-158 (b).

Grossberg, S. (Ed.), **The adaptive brain, I: Cognition, learning, reinforcement, and rhythm**. Amsterdam: Elsevier/North-Holland, 1987 (c).

Grossberg, S. (Ed.), **The adaptive brain, II: Vision, speech, language, and motor control**. Amsterdam: Elsevier/North-Holland, 1987 (d).

Grossberg, S., Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 1987, **11**, 23-63 (e).

Grossberg, S. and Kuperstein, M., **Neural dynamics of adaptive sensory-motor control: Ballistic eye movements**. Amsterdam: Elsevier/North-Holland, 1986.

Grossberg, S. and Levine, D., Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology*, 1975, **53**, 341-380.

Grossberg, S. and Pepe, J., Spiking threshold and overarousal effects in serial learning *Journal of Statistical Physics*, 1971, **3**, 95-125.

Hastings, S.P., The existence of periodic solutions to Nagumo's equation. *Quarterly Journal of Mathematics*, 1976, **27**, 123-134.

Hastings, S.P., Single and multiple pulse waves for the FitzHugh-Nagumo equations. *SIAM Journal of Applied Mathematics*, 1982, **42**, 247-260.

Hinton, G.E. and Anderson, J.A. (Eds.), **Parallel models of associative memory**. Hillsdale, NJ: Erlbaum, 1981.

Hodgkin, A.L., **The conduction of the nervous impulse**. Liverpool: Liverpool University, 1964.

Hodgson, J.E.P. (Ed.), **Oscillations in mathematical biology**. New York: Springer-Verlag, 1983.

Hopfield, J.J., Neuronal networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 1982, **79**, 2554-2558.

Hopfield, J.J., Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 1984, **81**, 3058-3092.

Kaczmarek, L.K. and Babloyantz, A., Spatiotemporal patterns in epileptic seizures. *Biological Cybernetics*, 1977, **26**, 199–208.

Kandel, E.R. and Schwartz, J.H., **Principles of neural science**. New York: Elsevier/North-Holland, 1981.

Katz, B., **Nerve, muscle, and synapse**. New York: McGraw Hill, 1966.

Kilmer, W.L., On some realistic constraints in prey-predator mathematics. *Journal of Theoretical Biology*, 1972, **36**, 9-22.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., Optimization by simulated annealing. IBM Thomas J. Watson Research Center, Yorktown Heights, New York, 1982.

Kosko, B., Bidirectional associative memories. Submitted for publication, 1987.

Kosko, B. and Guest, C., Optical bidirectional associative memories. *Society for Photo-optical and Instrumentation Engineering (SPIE) Proceedings: Image Understanding*, 1987, **758**, in press.

Levine, D.S., Neural population modeling and psychology: A review. *Mathematical Biosciences*, 1983, **66**, 1-86.

Lotka, A.J., **Elements of mathematical biology**. New York: Dover, 1956.

MacArthur, R.H., Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1970, **1**, 1-11.

Plonsey, R. and Fleming, D.G., **Bioelectric phenomena**. New York: McGraw-Hill, 1969.

Rumelhart, D.E. and McClelland, J.L. (Eds.), **Parallel distributed processing**, Vol. I. Cambridge, MA: MIT Press, 1986.

Sperling, G., Mathematical models of binocular vision. In S. Grossberg (Ed.), **Mathematical psychology and psychophysiology**. Providence, RI: American Mathematical Society, 1981.

Takeuchi, Y., Adachi, N., and Tokumaru, H., The stability of generalized Volterra equations. *Journal of Mathematical Analysis and Applications*, 1978, **62**, 453-473.

END
DATE
FILMED
6-1988
DTIC